DOCUMENT RESUME

ED 082 218 CS 200 702

AUTHOR Meyer, Bonnie J. F.
TITLE Identifying Variables in Prose.
PUB DATE May 73
NOTE 12p.; Paper presented at the Annual Meeting of the Eastern Psychological Assn. (44th, Washington, D. C., May 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Componential Analysis; *Content Analysis; Educational Research; *Learning Processes; Linguistics; *Memory; *Prose; Semantics; *Structural Analysis
IDENTIFIERS *Content Structure

ABSTRACT

This paper identifies and describes three distinct variables on which prose passages can differ, from a linguistic point of view. Each of these variables relates to the way in which the content of prose is organized to convey meaning to the reader. The first variable is the position of information in a structure representing a passage's organization. The second variable deals with the pattern of organization of a passage. The third deals with the amount of signaling given by the author to help the reader use the passage's organization. It is suggested that these three variables, which tap different aspects of the organization of information in prose, will be valuable tools for investigating how people learn and remember information from prose. (LI)

FILMED FROM BEST AVAILABLE COPY

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 082218

IDENTIFYING VARIABLES IN PROSE

Bonnie J. F. Meyer
Cornell University

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
Bonnie J. F. Meyer
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

Paper Presented at the
Eastern Psychological Association
44th Annual Meeting
Washington, D.C.
May, 1973

CS 200 702

Identifying Variables in Prose

Bonnie J. F. Meyer
Cornell University

The question of how people learn and remember information from complex written materials has long been of interest to psychologists. A comprehensive investigation of this question has been delayed due to the complexity of prose materials and the lack of skills necessary to control the variance of prose. A primary limitation of those studies which have used natural prose as the learning stimulus is that the results obtained from one passage cannot be generalized to another passage since the variables on which the passages are similar or different have not been specified. Attempts have been made in the past to categorize prose by assessing its readability through such measures as sentence length, vocabulary density and rarity, and idea density. These measures have had limited success since they deal with surface factors of the prose and are not concerned with the meaning of prose and the manner in which the content is organized to convey this meaning to the reader. Recent work by linguists (Fillmore, 1968; Longacre, 1968; Halliday, 1968; Frantz, 1970; Grimes, 1972) has dealt with meaning in prose and how relationships among ideas in prose build upon each other to convey a message. Relationships among ideas in language have been specified for use in analyzing and diagraming sentences and passages. Thus, useful frameworks have recently been provided from which to investigate organization in prose.

Utilizing concepts and methods from a linguistic framework, I have analyzed a number of passages in depth. From my analyses I have identified three distinct variables on which passages can differ. Each of these variables relates to the way in which prose is organized. The first variable is the position of information in a structure representing a passage's organization. The second deals with the pattern of organization of a passage. The third deals with the amount of signaling given by the author to help the reader use the passage's organization. These variables are useful for specifying what information is available in prose to be picked up by the reader.

This paper will describe these three variables and possibilities for their use in studies concerned with learning and memory from prose. Since each of the variables is related to the organization of information in a passage, the discussion of the variables will be preceeded by a brief explanation of the methodology employed to analyze a passage's organization. A more detailed description of the analysis procedures and examples of its use can be found in a paper which I recently presented at the Structural Learning Conference in Philadelphia (Meyer, 1973). The paper is available upon request.

The Content Structure

The methodology for prose analysis which I use yields hierarchically arranged tree structures. Nodes in these tree structures contain content words from the passage, and the lines among the nodes show spacially how the content is organized. In addition, labels are found in the tree structures which explicitly state and classify the relationships among the nodes. I call this hierarchically arranged structure of the passage's content the content structure. This type of structure has been previously called the logical structure and the semantic structure. These terms are easily confused with a structure in a person's memory. Therefore, to disambiguate these terms, a passage's structure will be referred to as the content structure since it shows the structure of the content of a passage.

To obtain the content structure of a passage I utilize Grimes' (1972) semantic grammar of propositions. Although I have made some changes in the use of Grimes' system, his basic tenants have been followed. In addition to the objectivity and descriptive powers provided by Grimes's system, I find it particularly appealing since my previous (Meyer, 1971) intuitive analyses of prose produced content structures equivalent to those obtained by the semantic grammar of propositions. People appear to have similar intuitive knowledge of the content structure of a passage as indicated by a 91.5% average agreement among the structures made intuitively by independent judges (Meyer, 1971).

The Semantic Grammar of Propositions

Rules of the Grammar: The semantic grammar of propositions consists of two rules, the Predicate Rule and the Argument Rule. These rules are represented symbolically in Figure 1.

Predicate Rule: $F \rightarrow p_1^* \quad A_0^*$

Argument Rule: $A \rightarrow (F \| i)$

Figure 1

The Predicate Rule explains that a proposition (F) is made up of one or more (*) predicates (p), or relations, and their arguments (A). A passage usually consists of one complex proposition. The Argument Rule shows that an argument (A) is a dummy symbol and can be replaced by another proposition (F), a referential index (i), or both. A referential index stands for the thing to which the author is referring. The terminal (p, i) and nonterminal (F, A) symbols in the Predicate and Argument Rules enable the grammar to be used to analyze prose of any complexity or length.

The predicates referred to are of two types: lexical predicates and rhetorical predicates. The lexical predicates are related to their arguments by specific semantic roles or cases, while the rhetorical predicates are not related to their arguments by these specified roles.

Lexical Predicates: They are related to their arguments by specific semantic roles. Lexical predicates are often verbs and their adjuncts that are actually present in a passage. The lexical predicates have particular roles that they must take as arguments, other roles that they may take as arguments, and still others which they cannot take as arguments. Several recent papers (Perfetti, 1972; Rumelhart, Lindsay and Norman, 1972; Kintch, 1972) have dealt with role relationships when explaining the use of Fillmore's (1968) case grammar. I use Grimes's role system which includes the roles, Agent, Patient, Experiencer, Instrument, Goal, Source, Noninstigative Cause, Range, Benefactive, Factitive and Essive. In addition, I have added Manner and Time to the list of roles. Table 1 gives a brief description of these roles.

Rhetorical Predicates: They are not related to their arguments by specified roles. Rhetorical Predicates are primarily responsible for giving prose its overall organization. They relate together lexical propositions (lexical predicates and their arguments) and also rhetorical propositions (rhetorical predicates and their arguments). Rhetorical predicates are a small number of explicit organizing relations in prose. There are three kinds of rhetorical predicates, Paratactic, Hypotactic, and Neutral. Paratactic rhetorical predicates have at least two arguments of equal weight; for example, a passage stating a problem and then the solution to the problem has a paratactic rhetorical

TABLE 1

Role Relationships

Agent	Animate instigator of an action
Experiencer	Living thing that perceives an action (examples: thing that knows, feels, hears thinks)
Noninstigative Cause	Inanimate cause of an action
Instrument	Something used inanimately to perform an action
Patient	Who or what is directly affected by an action
Benefactive	Someone or something on whom an action has a secondary effect, good or bad
Goal	Where (person, place or thing) an action is headed or where it ends up
Source	Where (person, place or thing) an action begins
Range	Place where an action is carried out
Time	Time in which an action is carried out
Factitive	Result of an action
Manner	Way an action is performed (examples: carefully, slowly)
Essive	Identifies an animate or inanimate thing with its properties. (Thing and properties are linked together with words like have and be [Langendoen, 1970])

predicate called Responce with the two equally weighted arguments, problem and solution. In contrast, the arguments of the hypotactic rhetorical predicates are not of equal weight. One argument is superordinate to the other arguments that describe or give further information about it. For example, a passage that states a problem and then gives more information about this problem has the hypotactic rhetorical predicate, Specific, where the statement of the problem is the superordinate argument and the details about the problem comprise the subordinate arguments. When diagraming the the content structure of a passage the arguments of the paratactic rhetorical predicates are located at the same level of nodes in the hierarchical structure, while the superordinate argument of the hypotactic predicate is placed at a level higher in the tree structure than the hypotactic predicate's other arguments. The third kind of rhetorical predicates, neutral rhetorical predicates, can take either a paratactic or a hypotactic form depending on the content of the arguments and the emphasis given to them by a passage's author. The rhetorical predicates that fall under either the paratactic, hypotactic or neutral classifications are listed in Table 2 along with a brief description of each. The content structure of a paragraph is presented in Figure 2.

Position of Information in the Content Structure

The first variable dealing with the organization of information in prose is position of information in the content structure. The 500 to 650 word passages which I have analyzed have had from seven to 14 levels of nodes in their hierarchically arranged content structures. The position variable deals with whether information is located at top level nodes in the content structure or lower level nodes. This position variable relates to what educators have referred to as the gist or main ideas of a passage. The main ideas of a passage are found at the top nodes of the content structure, while the details of a passage tend to be located at the lower nodes. In a previous study (Meyer and McConkie, 1973) it was found that information at the top nodes of the content structure was recalled better than information low in the structure. In that study the content of the information at high and low positions in the content structure was not controlled. Thus, the results could have been due to different types of information which may be located in high and low structural positions. In order to control for the content of information while investigating the effect of position in the content structure on what is remembered from prose, I wrote two passages each containing one identical paragraph. The passages were written so that the identical paragraph, the target paragraph, was high in the content structure of one passage and low in the structure of the other passage. In addition, each passage was of equal length and the target paragraph came at the same serial location in both passages. I have written three sets of materials of this type. Preliminary analysis shows that the target paragraph is recalled over twice as well when it is high in a passage's content structure than when it is low in the structure.

These findings indicate that the organization in prose as represented by the content structure influences a reader's organization in memory of the passage's information, and thus, his subsequent recall. In order to investigate what information is in a reader's memory from high and low structural positions, I had subjects free recall a passage and then recall the target paragraph portion of the passage with cues. The cues were the content words from the content structure of the target paragraph. Thus, in the cued recall condition a subject's task was to write a paragraph providing the correct interrelationships among the cued content words. Examination of a proportion of this data shows that the target paragraph was recalled better in the cued recall condition than the free recall condition for all subjects, but that the cued recall condition facilitated recall to a much greater extent for those subjects who read the target paragraph in the

TABLE 2

Rhetorical Predicates

Paratactic Rhetorical Predicates	Description
Alternative	Equal weighted alternative options
Response	Equal weighted Question(s) and Answer(s), Remark and Reply, or Problem(s) and Solution(s)

Hypotactic Rhetorical Predicates	Description
Attribution	Describes qualities of person, place or thing
Equivalent	Restates same information in a different way
Specific	Gives more specific information about something that was stated in a more general manner
Explanation	Previously stated information is explained in a more abstract manner (for example: relating the information to a general principle) or more concrete manner
Evidence	Evidence through perception of a situation to support some idea
Analogy	Analogy given to support an idea
Setting Location	Gives location of setting in which information being related occurs (used particularly in narratives)
Setting Time	Gives time of setting in which information being related occurs (often in narratives)
Setting Trajectory	Gives changing background of location and time that occurs in a narrative when characters travel through various places
Representative Identification	Singles out one element of a group and makes it stand for the group as a whole
Replacement Identification	One thing standing for something else
Constituency Identification	Identifies a part in relation to some whole

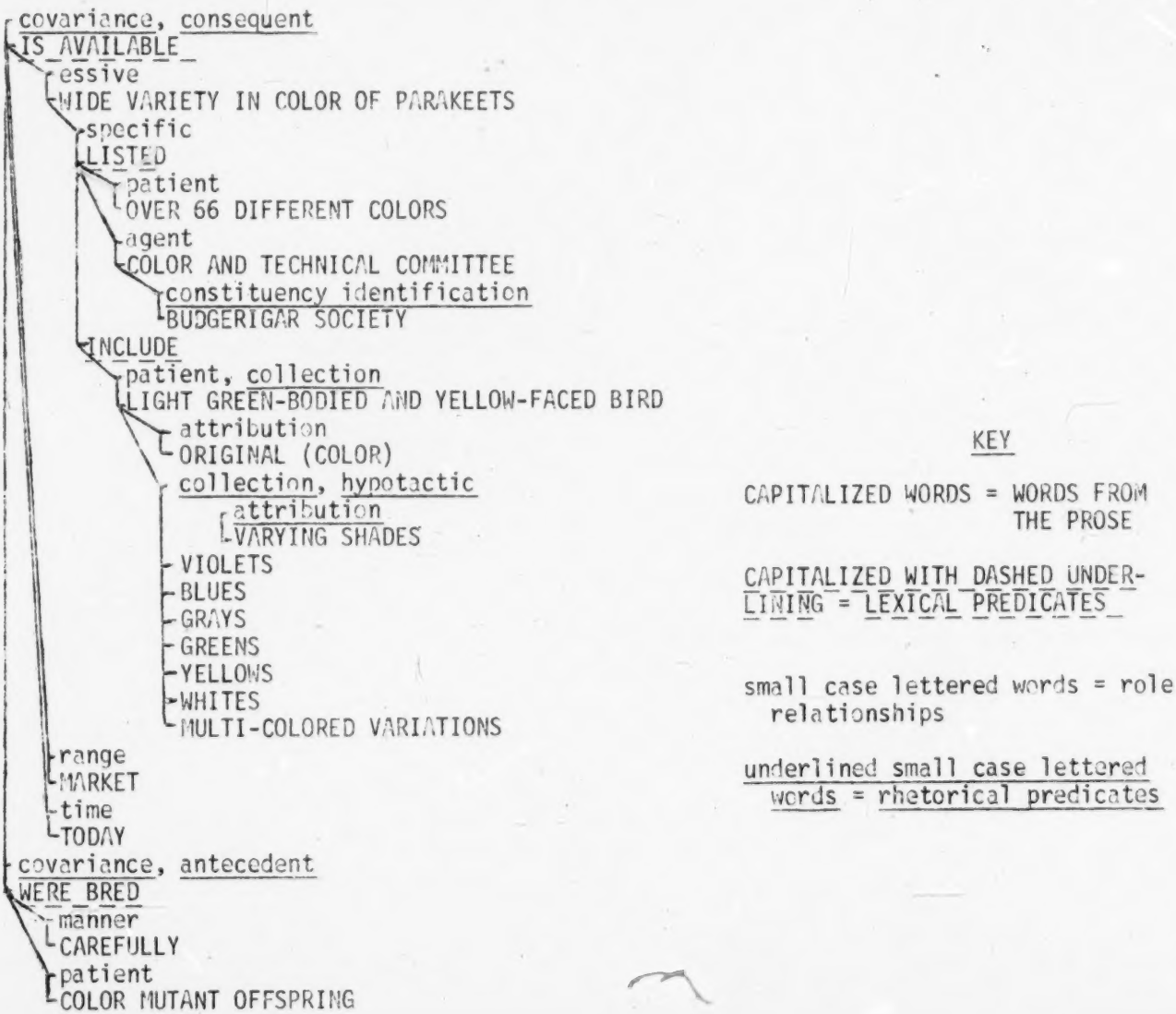
Neutral Rhetorical Predicates	Description
Collection	List of elements related in some manner
Covariance	Relation often referred to as condition, result, or purpose with one argument serving as the Antecedent and the other as the Consequent, result of the antecedent
Adversative	Relates what did not happen to what did happen

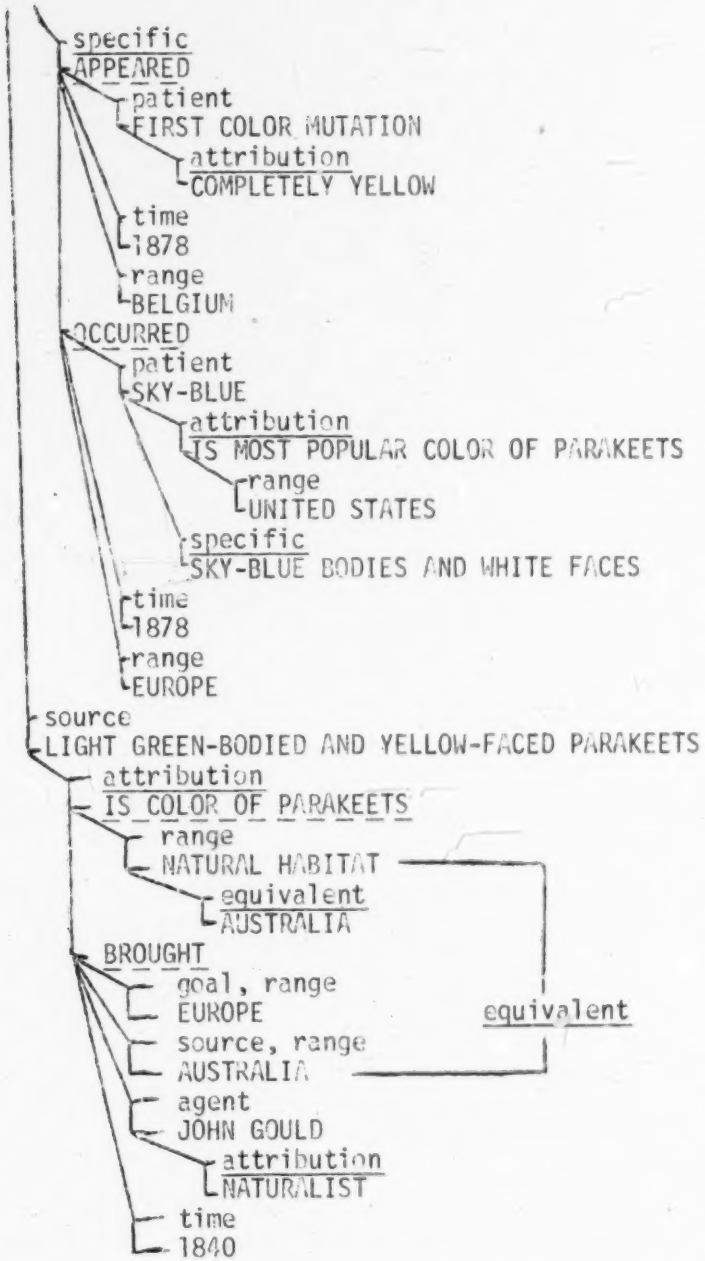
Figure 2

Content Structure of the Parakeet Paragraph

Parakeet Paragraph

The wide variety in color of parakeets that are available on the market today resulted from careful breeding of the color mutant offspring of light green-bodied and yellow-faced parakeets. The light green body and yellow face color combination is the color of parakeets in their natural habitat, Australia. The first living parakeets were brought to Europe from Australia by John Gould, a naturalist, in 1840. The first color mutation appeared in 1872 in Belgium; these birds were completely yellow. The most popular color of parakeets in the United States is sky-blue. These birds have sky-blue bodies and white faces; this color mutation occurred in 1878 in Europe. There are over 66 different colors of parakeets listed by the Color and Technical Committee of the Budgerigar Society. In addition to the original light green-bodied and yellow-faced birds, colors of parakeets include varying shades of violets, blues, grays, greens, yellows, whites and multi-colored variations.




specific
APPEARED
patient
FIRST COLOR MUTATION
attribution
COMPLETELY YELLOW
time
1878
range
BELGIUM
OCCURRED
patient
SKY-BLUE
attribution
IS MOST POPULAR COLOR OF PARAKEETS
range
UNITED STATES
specific
SKY-BLUE BODIES AND WHITE FACES
time
1878
range
EUROPE
source
LIGHT GREEN-BODIED AND YELLOW-FACED PARAKEETS
attribution
IS COLOR OF PARAKEETS
range
NATURAL HABITAT
equivalent
AUSTRALIA
BROUGHT
goal, range
EUROPE
source, range
AUSTRALIA
agent
JOHN GOULD
attribution
NATURALIST
time
1840
equivalent

low content structure position. Although these subjects recall increased from 50 to 100% in the cued recall condition, their cued recall performance was still poorer than either the free recall performance or cued recall performance of the subjects who read the target paragraph in the high content structure position. This data suggests that information low in the content structure is less often stored in memory than information high in the content structure, and what is stored is less available for retrieval by the reader.

Pattern of the Content Structure

The second variable is the pattern of a passage's content structure. Removing the content words from the nodes in the content structure leaves a pattern of the structure; remaining in the structure are the role relationships and rhetorical predicates which show how the content was organized. I have written two passages with identical structures; that is, the passages have the same patterns with identical role relationships and rhetorical predicates, but different content. Recall protocols from these passages will be examined from similarities due to identical structure patterns and differences due to content.

Passages differ in the pattern of their top level content structures. This pattern at the top of the content structure gives a passage its overall organization. The pattern of the top level of the content structure can be classified according to the type of rhetorical predicates, or less commonly role relationships, that occur at the top of the content structure. A common top level pattern found both in fairy tales and scientific articles is the rhetorical predicate, Response of the Problem - Solution variety. Other top level patterns include Alternative, Event - Source - Attribution and/or Specification of the Source, Event - Attribution, Event - Evidence, Event - Analogy, Collection with time sequencing (using Litteral's (1971) time topology; Collections of this type are often found in narratives), Covariance (Antecedent - Consequent, Event - Result, Event - Purpose), and Adversative. Figure 3 depicts schematizations of some top level structures.

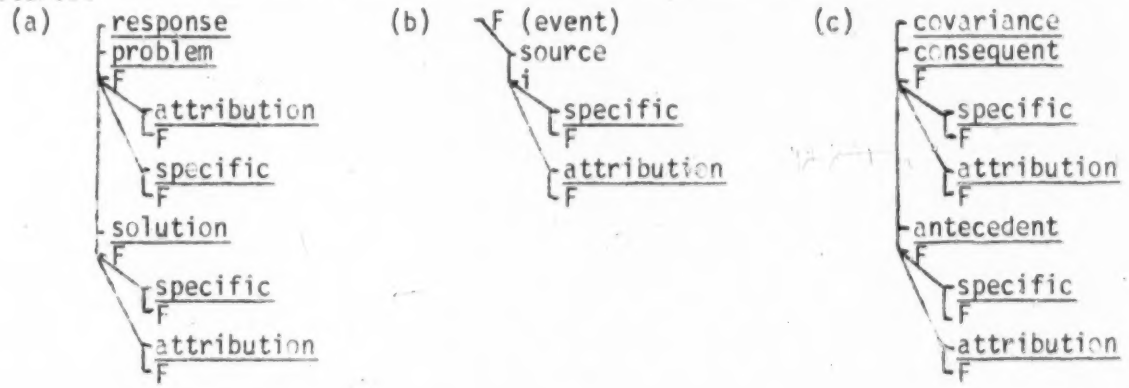


Figure 3

It has been asserted that authors through the ages who wanted certain information to be remembered put that information in the Problem - Solution pattern. The efficacy of certain top level structures over others in facilitating memory can be tested due to the fact that passages can be written which place the same topic in different top level patterns.

An example of placing the same topic in different top level structures is the topic the therapeutic community (form of mental health treatment) which could be placed in any one of the three top level patterns in Figure 3. A passage which I analyzed (Meyer, 1971) dealt with this topic using the top level pattern (b) in Figure 3. The passage's first paragraph dealt with how the treatment of mental patients had changed in the last 20 years. The establishment of therapeutic communities was given as the third source of the shifts in the treatment of the mentally ill. The remainder of the passage dealt with the therapeutic community. The author's purpose in writing the article appeared to be describing and selling the therapeutic community to his readers, rather than discussing a number of changes in the treatment of mental illness over the last 20 years as indicated by his first paragraph. Subjects' recall of this passage was best for information in the first paragraph (event and source in Figure 3 -b), and poor for the description of the therapeutic community. This topic could be rewritten in top level pattern a or c and recall protocols from the three forms of the topic could be examined. In pattern a, the Problem - Solution top level pattern, the problem would consist of trouble incurred in treating mental illness and the solution would be the therapeutic community. In c, the Covariance top level pattern, the consequent would be the improvement in the treatment of mental illness and the antecedent would be the therapeutic community.

Amount of Signaling Given to the Reader

The third variable deals with the amount of signaling given by the author to help the reader identify and use the content structure of the passage. This signaling variable is similar to the types of relationships referred to by Halliday as theme and by Grimes as staging. The information in prose which I am referring to as signaling does not appear in the content structure since it is extra information used to guide the reader to the author's organization of his ideas. Information defined as signaling information is of the following types:

1. Prematurely given information: Information in the content structure of the passage state either with the same words or paraphrased wording toward the beginning of a passage or paragraph as it is stated later in the text. The later occurrence of the information is where the information belongs in relation to the other information in the content structure of the prose.

   often seen in: titles, introductory sentences of passages and paragraphs

2. Summary statements: This type of signaling is similar to that mentioned in point one except that the information is not given prematurely. Instead, the same words or paraphrased wording for information already presented and located in the content structure are stated again at the end of a paragraph or a passage.

   often seen in: summary statements at the end of paragraphs or passages

3. Pointer words: Words that explicitly tell the reader that a certain idea is important.

   example: an important point is _____

4. Specification of aspects of the content structure: Explicitly stated rhetorical predicates or other information contained in the content structure.

examples: The problems are _____ and _____. The solution to these problems is _____.

Three factors caused _____. One, _____. Two, _____. Three, _____.

The content structure of two passages can be held constant while varying the presence of signaling. Some passages in their natural form have more signaling provided by their authors than others. I reduced one passage from 638 words to 494 words by eliminating this signaling aspect of the prose, while another passage of about the same length was only reduced by 14 words. Signaling phrases and sentences can be generated from the content structure of a passage. Thus, experimental conditions can be designed to test the influence of this variable of signaling. I am in the process of analyzing recall protocols of subjects who read passages with or without signaling. This variable appears to influence the amount and type of information remembered.

In conclusion, these three variables, which tap different aspects of the organization of information in prose, will be valuable tools for investigating how people learn and remember information from prose. A person's organization of the learning stimulus has been shown to be important in determining what he will remember from lists of words (Tulving, 1968; Mandler, 1967). Preliminary findings suggest that the organization in a passage affects a reader's selection and organization of information from the passage for storage in his memory, and thus, it affects his subsequent recall of the passage. These three variables will assist in the specification of the relationships among the organization of information in a passage, in a reader's memory of the passage, and in his recall protocol.

References

Fillmore, C. J. The case for case. In E. Bach and R. Harms (Eds.), Universals in Linguistic Theory. New York: Holt, Rhinehart & Winston, 1968, 1-81.

Frantz, D. G. Toward a generative grammar of Blackfoot. Doctoral dissertation, University of Alberta, 1970.

Grimes, J. E. The Thread of Discourse. Ithaca, Cornell University, 1972.

Halliday, M.A.K. Notes on transitivity and theme in English, Part 3. Journal of Linguistics, 4, 179-215.

Kintch, W. Notes on the structure of semantic memory. In E. Tulving and W. Donaldson (Eds), Organization of Memory, New York: Academic Press, 1972, 247-308.

Langendoen, D. T. Essentials of English Grammar. New York: Holt, Rinehart & Winston, 1970.

Litteral, R. Rhetorical Predicates and Time Topology in Anggor. Summer Institute of Linguistics, manuscript, 1971.

Longacre, R. E. Discourse, Paragraph, and Sentence Structure in Selected Philippine Languages. Vol. 1, Santa Anna: Summer Institute of Linguistics, 1968.

Mandler, G. Organization and Memory. In K. W. Spence and J. T. Spence (Eds.), The Psychology of Learning and Motivation Advances in Research and Theory. New York: Academic Press, 1967, Vol. 1, 328-372.

Meyer, B. J. F. Idea Units Recalled from Prose in Relation to their Position in the Logical Structure, Importance, Stability and Order in the Passage. Cornell University, M. S. thesis, 1971.

Meyer, B. J. F. Structure of Prose: Identification and Effects on Memory. Paper presented at Fourth Annual Invitational, Interdisciplinary Meeting on Structural Learning. Philadelphia, Pennsylvania. April, 1973.

Meyer, B. J. F. and McConkie, G. W. What is recalled after hearing a passage? Journal of Educational Psychology, 1973, in press.

Perfetti, C. A. Psychosemantics: Some cognitive aspects of structural meaning. Psychological Bulletin, 78, 1972, 241-259.

Rumelhart, D. E., Lindsay, P. H., and Norman, D. A. A process model for long-term memory. In E. Tulving and W. Donaldson (Eds.), Organization of Memory, New York: Academic Press, 1972, 195-246.

Tulving, E. Theoretical issues in free recall. In T. R. Dixon and D. L. Horton (Eds.), Verbal Behavior and General Behavior Theory. New Jersey: Prentice Hall, 1968.